Math 365 / Comp 365: Exam 2

Instructions:

- This portion of the exam should be added to the back of the written part, and you should submit a single, stapled copy.
- You must also submit this file in Moodle.
- Please fill in your name in the appropriate place below.
- Please leave intact the leading blocks which load the Matrix package, the `365FUnctions.r` file, and the data for problem 5.
- Please read carefully the instructions for each problem.
- Please preserve all of the section headings for solutions, and insert your solutions in the appropriate places. If you used R to solve any of the written problems, create careful section headings like I have done here.
- You are certainly allowed to pull code from things we've used/developed in this course (e.g., any code I gave you as part of in-class activities).

Yuxiang Chen

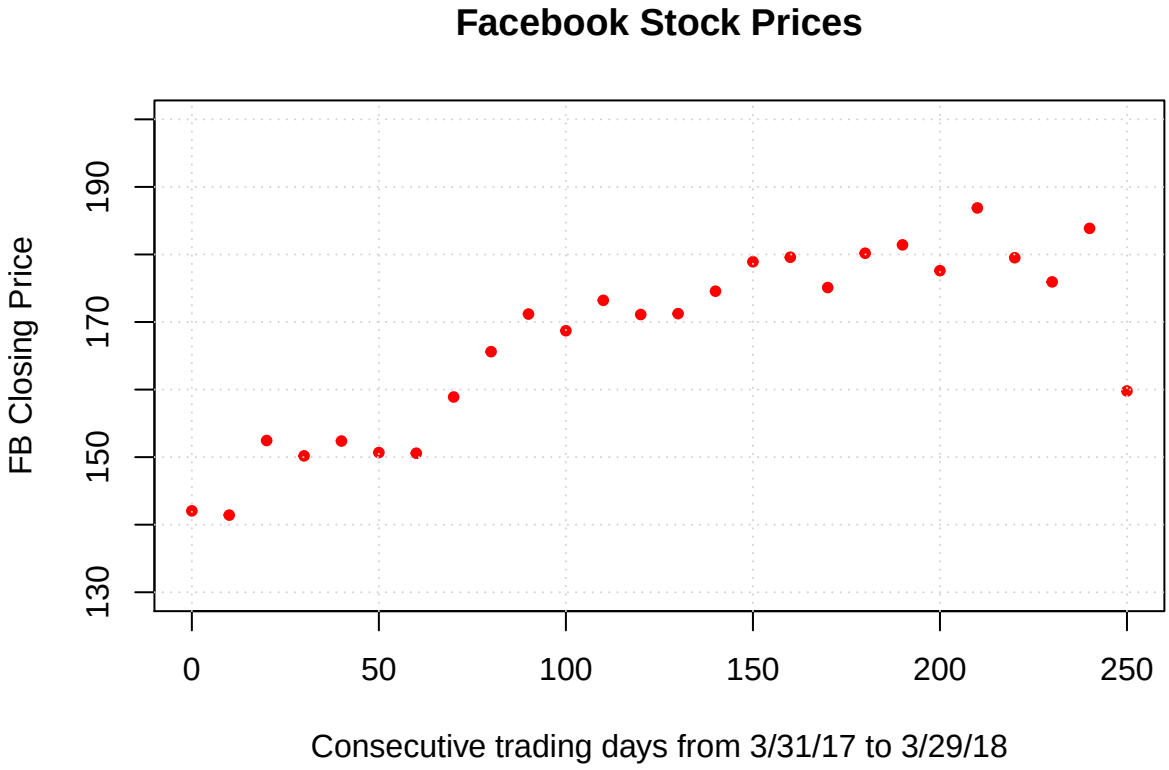
```
require(Matrix)
require(mosaicData)
source("https://drive.google.com/uc?export=download&id=10dNH3VbvxS8Z3OHjP4i9gRbtsf91VVBb") #365function
source("https://drive.google.com/uc?export=download&id=1et61nwz2dQzJ6TiJlNpGNx5vtVrqHMmE") #facebook.pr
```

Problem 5

The stock price of Facebook has been in the news a lot lately, so I was interested to see how it has fluctuated throughout the last year. I accidentally only downloaded the closing prices at every 10^{th} trading day (roughly every two weeks) for the past 251 trading days (one trading year from 3/31/2017 to 3/29/2018). Let's examine three different ways to fill in the missing data.

For now, here is a plot of the data we have. Note: the prices are stored in `samp.prices`, which I have loaded for you.

```
days = 0:250
samp.days = seq(0,250,by=10) # we only have data on these days
plot(samp.days,samp.prices,pch=20,cex=.5,ylim=c(130,200),
     ylab = "FB Closing Price",
     xlab = "Consecutive trading days from 3/31/17 to 3/29/18",
     main = "Facebook Stock Prices")
points(samp.days,samp.prices,pch=20,cex=1,col='red')
grid()
```



Here is the function we are going to use to plot the different approximation models:

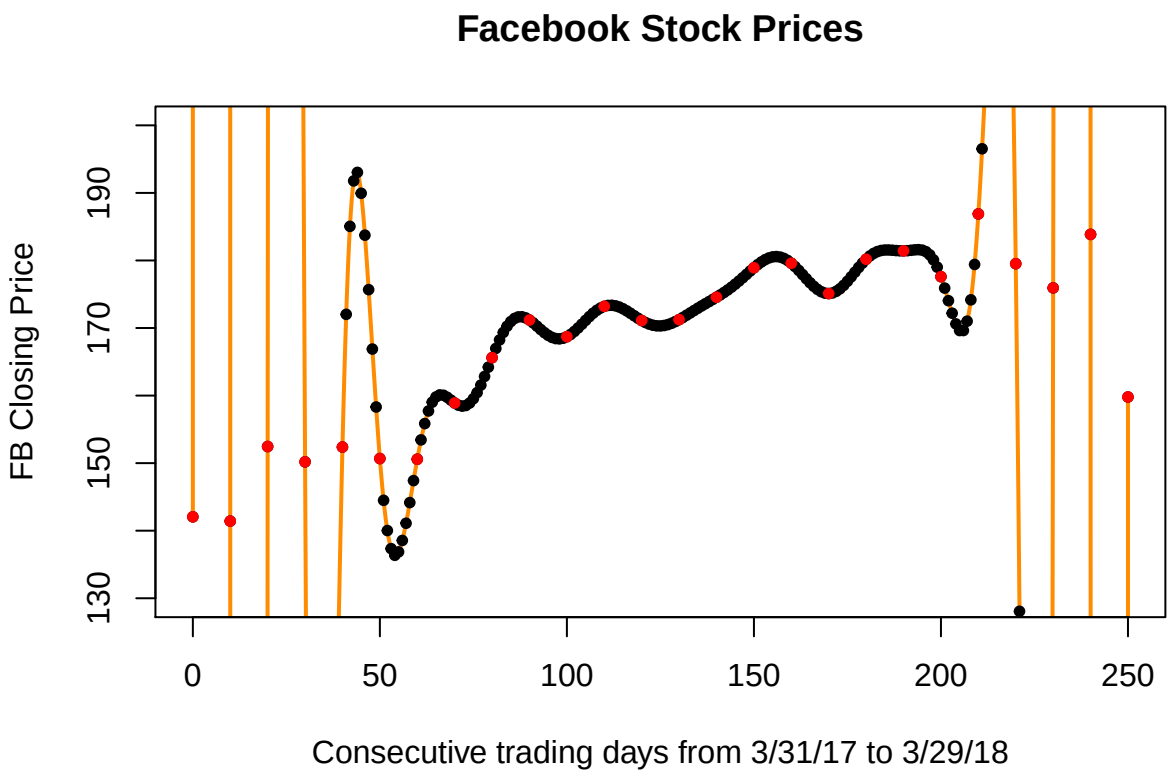
```
plotApproximations=function(f){
  tt = seq(0,250,length=1001)
  plot(tt,f(tt),col="DarkOrange",type='l',lwd=2,ylim=c(130,200),
       ylab = "FB Closing Price",
       xlab = "Consecutive trading days from 3/31/17 to 3/29/18",
       main = "Facebook Stock Prices")
  points(days,f(days),pch=20,cex=1,col='black') # The points guessed by the interpolating/regressing fu
  points(samp.days,samp.prices,pch=20,cex=1,col='red') # The sample points we have from every 10th day
}
```

Part (a):

Fit a degree 25 interpolating polynomial to this data. Call the function that evaluates the polynomial `p`. Write your code in the block below and uncomment the last line there to display your answer.

Part (a) solution

```
coeff = NewtonDD(samp.days,samp.prices)
p = function(t){Horner(coeff,t,samp.days)}
plotApproximations(p)
```

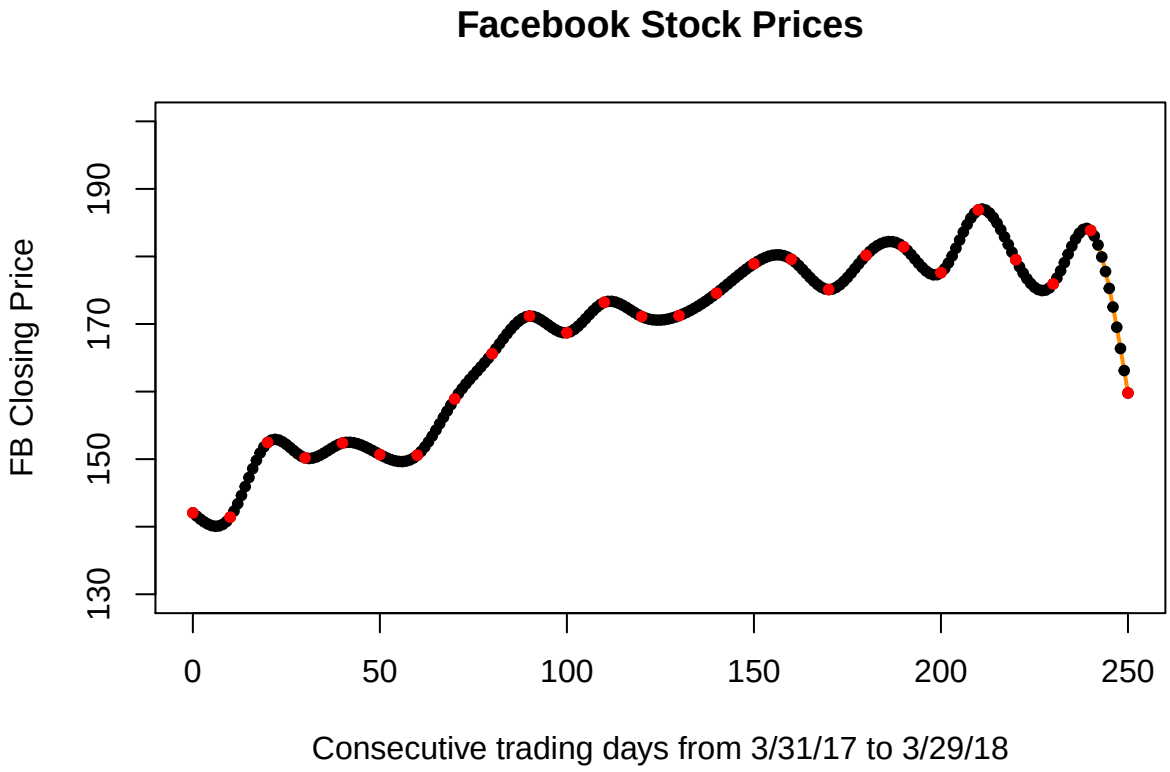


Part (b):

Compute a natural cubic spline passing through the 26 stock prices. Call the function that evaluates the spline `s`. Write your code in the block below and uncomment the last line there to display your answer.

Part (b) solution

```
s = splinefun(samp.days,samp.prices,method = "natural")
plotApproximations(s)
```



Part (c):

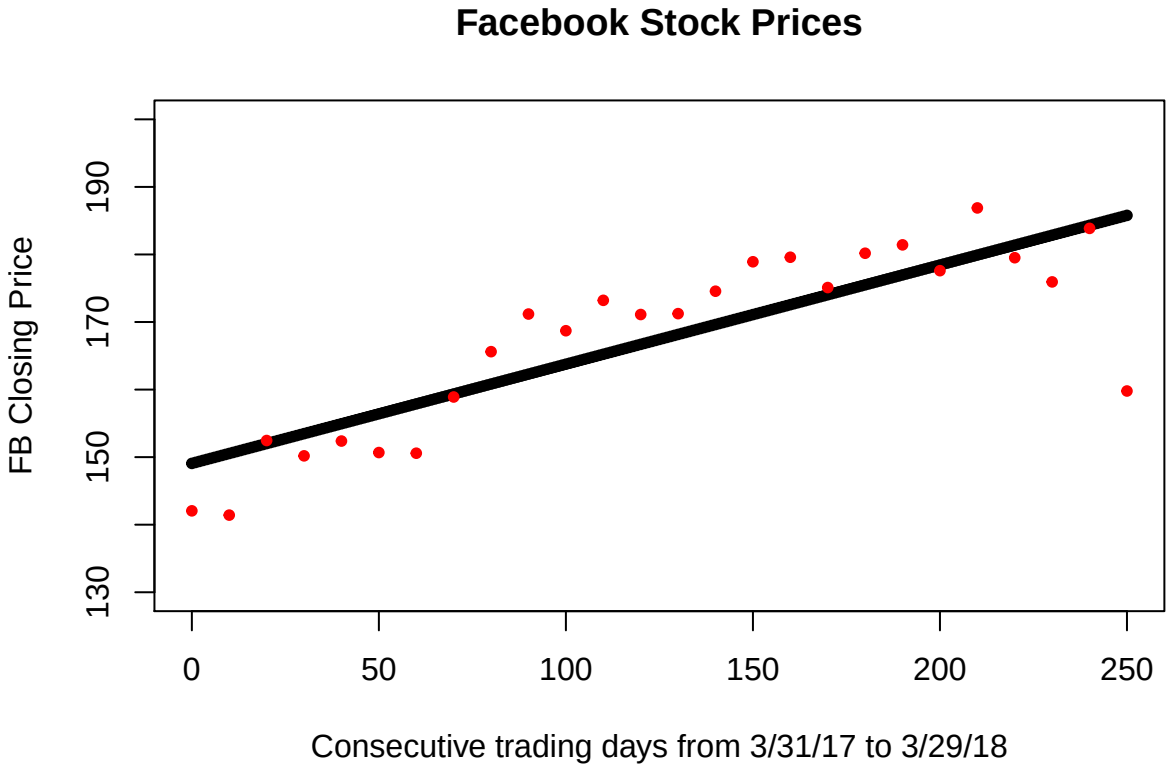
Solve the least squares problem to fit a regression line through the data. Do not use any statistical modeling commands like `lm` for this part of the problem; rather, you may want to (though you do not have to) use the function `qr.solve()` to solve the least squares problem. Write your code in the block below and uncomment the last line there to display your answer.

Part (c) solution

```
A = cbind(rep(1,26),samp.days)
qr.solve(A,samp.prices)
```

```
##                samp.days
## 149.0720527   0.1468359
```

```
r=function(t){149.0720527+0.1468359*t}
plotApproximations(r)
```



Part (d):

Aha! I found the full data after all. Turns out it was also right here this entire time, and it is stored as `all.prices`. All the days are stored in `days`. For each of the three methods above, compute the squared 2-norm of the vector of errors between the true values of the stock price and your estimated error. Do this by completing the first three lines in the code block below, and leave the following three lines to print the answers. Finally, on the last line, fill in the proper information to order the three methods from best to worst in terms of how they performed with the squared 2-norm.

Part (d) solution

```
x=seq(0,250)
poly.err <- norm(p(x)-all.prices,"2") #REPLACE 0 WITH YOUR CODE
spline.err <- norm(s(x)-all.prices,"2") #REPLACE 0 WITH YOUR CODE
reg.err <- norm(r(x)-all.prices,"2") #REPLACE 0 WITH YOUR CODE
print(paste("Squared 2-norm of error for polynomial interpolation =",poly.err))
```

```
## [1] "Squared 2-norm of error for polynomial interpolation = 340926.674499865"
```

```
print(paste("Squared 2-norm of error for spline interpolation =",spline.err))
```

```
## [1] "Squared 2-norm of error for spline interpolation = 58.3603463531277"
```

```
print(paste("Squared 2-norm of error for linear regression =",reg.err))
```

```
## [1] "Squared 2-norm of error for linear regression = 116.34732639963"
```

```
print(c("BEST METHOD:spline interpolation","NEXT BEST METHOD: linear regression","WORST METHOD: polynomi
```

```
## [1] "BEST METHOD:spline interpolation"
## [2] "NEXT BEST METHOD: linear regression"
## [3] "WORST METHOD: polynomial interpolation"
```

Problem 6

This question reveals the math behind an important machine learning algorithm called 'Ridge Regression'.

Part a:

We will consider a dataset recording the average total SAT score of each state.

```
SAT = mosaicData::SAT[,c(2,3,4,5,8)]
SAT.S = scale(SAT)
```

I have extracted columns from the data to contain 4 input variables (expenditure per pupil in a state `expend`, pupil/teacher ratio `ratio`, teacher's average salary `salary`, and percent of eligible students taking SAT `frac`), and the 5th column is the predictor variable (total SAT score `sat`). I have also standardized the data, and called this scaled data `SAT.S`. (Check out both `SAT` and `SAT.S`. We would like to model the (scaled) SAT score as a linear function of each of the (scaled) inputs:

$$\texttt{sat} = x_1\texttt{expend} + x_2\texttt{ratio} + x_3\texttt{salary} + x_4\texttt{frac}$$

Part a(i) Set up a system of equations $Ax = b$ for this model, where the x represents the unknown coefficients of your model. Explain why your system is as you have determined. Note: this will be an overdetermined system.

Part a(i): Solution

```
A=cbind(SAT.S[,1],SAT.S[,2],SAT.S[,3],SAT.S[,4]) #Input the appropriate matrix here
b=SAT.S[,5] #Input the appropriate vector here
```

Part a(ii)

Use the normal equations to solve for the least squares solution of your system.

Part a(ii): Solution

```
#Insert code to solve for the least squares solution, call it sol
sol = qr.solve(A,b)
#Uncomment below
print(sol)
```

```
## [1]  0.08128318 -0.10977993  0.13006185 -1.03889785
```

Part b:

To set up the Tikhonov regularization, we form the optimization problem

$$\min_{x\in\mathbb{R}^n}\left\{\frac{1}{2}||Ax-b||_2^2+\frac{\lambda}{2}||x||_2^2\right\}, \tag{1}$$

to solve for x, where λ is a positive constant that penalizes large magnitudes of x, the coefficients of the model. Notice if $\lambda = 0$, the solution to the ridge regression problem and the least squares solution will be the same.

Part b(i)

Using the same 3 parts as in HW 6 problem 3, explain why solving the above optimization problem can be reduced to solving the equation $(A^TA + \lambda I)x = A^Tb$ for x.

Part b(i): Solution

I encourage you to hand write this part out on paper.

Part b(ii)

Write a function for the ridge regression `ridge.reg` that takes in the parameters `A, b, lambda`. Find the ridge regression result when λ is 10. ### Part b(ii): Solution

```
ridge.reg = function(A,b,lambda){
  # Solve the system in part b(i) and store the solution in x
  # Insert your code here
  A1 = t(A)%*%A+lambda*diag(nrow(t(A)%*%A))
  b1 = t(A)%*%b
  x = solve(A1, b1)
  return(x)
}
r = ridge.reg(A,b,50) #Uncomment and run once function is written
r
```

```
##              [,1]
## [1,] -0.037739591
## [2,] -0.009494328
## [3,] -0.077758727
## [4,] -0.405272538
```

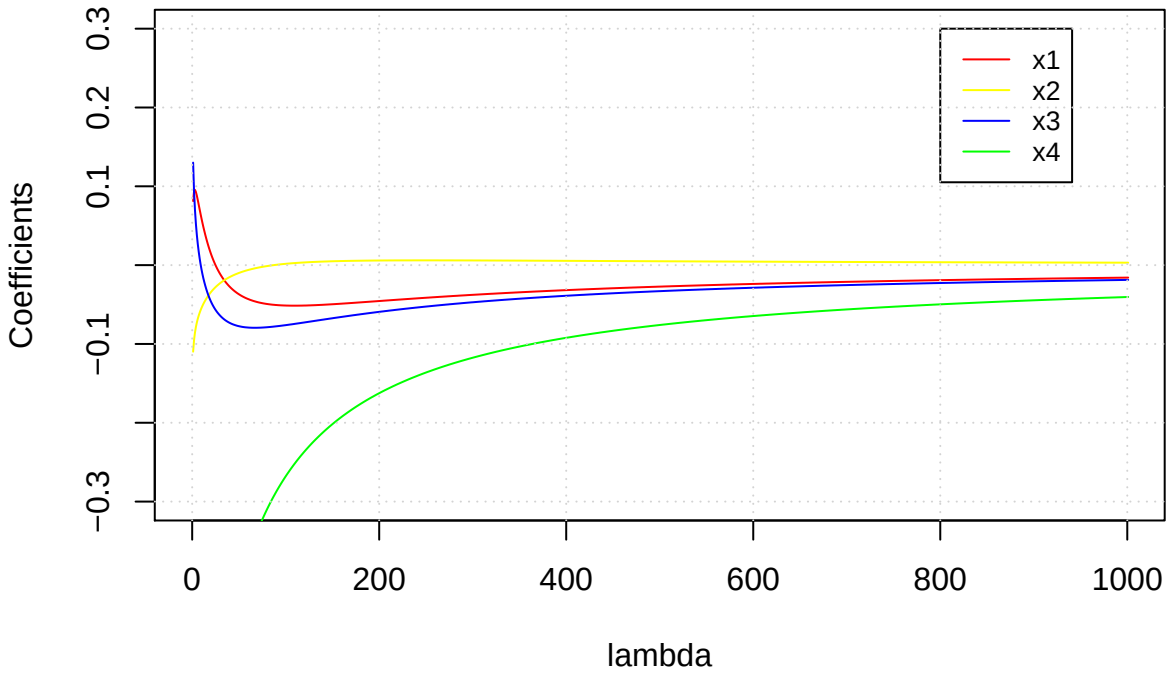
Part b(iii)

Plot the trend of each coefficient for λ ranging from 0 to 1000 (integer values). What do you observe (i.e. how does λ affect each coefficient)?

Part b(iii) Solution

```
#Insert code here to compute result, which should be a 4 x 1001 matrix for each of the coefficients
result=matrix(, nrow = 4, ncol = 1001)
for(i in 1:1001){
  result[,i]=ridge.reg(A,b,i-1)

}
##Uncomment code below to plot
plot(result[1,],type = 'l',xlim=c(0,1000),ylim = c(-0.3,0.3),col = "red",xlab="lambda",ylab="Coefficien
lines(result[2,],col = "yellow")
lines(result[3,],col = "blue")
lines(result[4,],col = "green")
legend(800,0.3,legend=c("x1","x2","x3","x4"),col=c("red","yellow","blue","green"),lty=1,cex=0.8)
grid()
```



As lambda grows larger, all coefficients become smaller and closer to 0.

Part b(iv)

There is a balancing act in trying to solve this optimization problem. The $\frac{1}{2}||Ax-b||_2^2$ term wants to minimize the squared residual error, while the $\frac{\lambda}{2}||x||_2^2$ term wants to minimize the magnitude of x. We would like to determine the λ that leads to the lowest minimum of the sum of both terms. One way to do this is a common method called "K-fold cross validation". Below is an outline of how to perform a K-fold cross validation:

- Choose a value K and a λ. We will fix $K = 10$, and vary λ from the integers 0 to 100.
- Divide the data into K equal parts. For our state SAT data, we separate the 50 states into 10 groups g_n where $n = 1, 2, 3, \ldots, 10$, each having 5 data points.
- For each $\ell = 1, 2, 3, \ldots, 10$, 'leave out' group g_ℓ in your data and fit the ridge regression model to all the other groups g_j where $j \neq \ell$ using parameter λ. Create the model using the coefficients that you find from the ridge regression, and use this model to predict a value for all the data points you left out in g_ℓ. Calculate the residual sum of squares $\sum_{i=1}^{5}(predict_i - b_i)^2$ where b_i is the actual scaled SAT score of the data points in g_ℓ.
- Now, you will have K residual sum of squares values, one for each 'left out' group. Compute the mean of these K values. This is called the cross validation value $CV(\lambda)$.
- Repeat the process with a different λ.
- The 'best' λ is the one with the lowest $CV(\lambda)$.

Implement $K = 10$-fold cross validation to find the best (integer) λ between 1 and 100.

Part b(iv) Solution

```
K = 10
BestLambda = 0 #Update this to be the optimal choice
bestError = 1000000 #Update this to the optimal choice (want it as close to 0 as possible)
for (lambda in 1:100){
  errorTotal = 0
  for (i in 1:K){
    ### Insert your code here
```

```
    AleftOut = A[((5*(i-1)+1):(5*i)),]
    bleftOut = b[((5*(i-1)+1):(5*i))]
    Arest =    A[-((5*(i-1)+1):(5*i)),]
    brest =    b[-((5*(i-1)+1):(5*i))]
    coeffs=ridge.reg(Arest,brest,lambda)
    m = function(A){A%*%coeffs}
    bpredicted=m(AleftOut)
    r1 = bpredicted-bleftOut
    error =(norm(r1,"2"))^2  #t(r1)%*%r1
    errorTotal=errorTotal+error
  }

  meanError = errorTotal/10
  #print(list(error,errorTotal,meanError))
  if(meanError<bestError){
    bestError = meanError
    BestLambda = lambda
  }
}
print(list(lambda = BestLambda,error = bestError))
```

```
## $lambda
## [1] 1
##
## $error
## [1] 1.052612
```

Part c

Observe two items

- $||x_a^*||_2^2 \geq ||x_b^*||_2^2$ where x_a^* corresponds to the optimal least squares solution in part (a) and x_b^* corresponds to the optimal ridge regression solution in part (b) using the best λ.
- $||Ax_a^* - b||_2^2 \leq ||Ax_b^* - b||_2^2$

Compute each of these quantities, and explain why this makes sense.

Part c Solution

```
A3 = t(A)%*%A+diag(4)
b3 = t(A)%*%b
sol2 = solve(A3,b3)
```

```
#Update each of these four values
norm_xa=norm(sol,"2")^2
norm_xb=norm(sol2,"2")^2
residual_xa=norm(A%*%sol-b,"2")^2
residual_xb=norm(A%*%sol2-b,"2")^2

print(paste("2-norm squared of least squares solution=",norm_xa))
```

```
## [1] "2-norm squared of least squares solution= 1.1148834202045"
```

```
print(paste("2-norm squared of ridge regression solution =",norm_xb))
```

```
## [1] "2-norm squared of ridge regression solution = 1.02677876303482"
```

```
print(paste("2-norm squared of residual for least squares solution =",residual_xa))
```

```
## [1] "2-norm squared of residual for least squares solution = 8.59644509902496"
```

```
print(paste("2-norm squared of residual for ridge regression solution =",residual_xb))
```

```
## [1] "2-norm squared of residual for ridge regression solution = 8.63888596209162"
```

Since the ridge regression is designed to minimize the magnitudes of x. Therefore the first bullet point makes sense. However,since the term $\frac{\lambda}{2}||x||_2^2$ deviates the R square from the least square solution, the R square for part a will be smaller than the r square for prt b.